WallTweet: A Knowledge Ecosystem for Supporting Situation Awareness

Teresa Onorati, Petra Isenberg, Anastasia Bezerianos, Emmanuel Pietriga, Paloma Diaz

WallTweet: A Knowledge Ecosystem for Supporting Situation Awareness

Teresa Onorati
Universidad Carlos III
Madrid, Spain
tonorati@inf.uc3m.es

Petra Isenberg
INRIA
Paris, France
petra.isenberg@inria.fr

Anastasia Bezerianos
U.Paris-Sud & CNRS, INRIA
Paris, France
anastasia.bezerianos@lri.fr

Emmanuel Pietriga
INRIA & INRIA Chile
Paris, France & Santiago, Chile
emmanuel.pietriga@inria.fr

Paloma Diaz
Universidad Carlos III
Madrid, Spain
pdp@inf.uc3m.es

ABSTRACT
We present WallTweet, a tweet visualization designed for wall displays and aimed at improving the situation awareness of users monitoring a crisis event utilizing tweets. Tweets are an important source of information during large-scale events, like tornados or terrorist attacks. Citizens in affected areas are often direct witnesses of the situation, and can be aware of certain details useful to, e.g., news channels and emergency response organizations. Yet, tweets are hard to visualize and put in a geographical context: large quantities of tweets get sent in a short period, that vary greatly in content and relevance with respect to the crisis at hand. Our visualization tool is currently a work in progress: it addresses these challenges by performing a semantic analysis of the tweets' content and displaying them on a ultra-high-resolution wall display. The goal of our tool is to create an inclusive experience that enhances users' situation awareness during a crisis event, by displaying geo-referenced tweets in detail, embedded into the more global geographic context of the event.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Tweet visualization; wall displays; crisis monitoring.

INTRODUCTION
We are currently working on a geographic visualization tool called WallTweet that visualizes a knowledge ecosystem built from the analysis of information collected from Twitter during a large-scale event. WallTweet's goal is to improve the situation awareness of people involved in crisis monitoring through integrated details and overview visualizations for ultra-high-resolution wall displays (see Figure 1).

Figure 1. Clusters of tweets labeled by the most relevant words.

The datasets underlying WallTweet are tweets (brief textual messages) collected from Twitter during a crisis event and then analyzed. These datasets are becoming more prevalent as the wide availability of modern smartphones with photo, GPS, and video capabilities has led citizens to actively report on crisis events [15]. It is now very common that, when an emergency occurs, citizens start to share information about the situation, not only as witnesses, but also driven by curiosity. For example, YouTube lists more than 1 million amateur videos for the search terms *tsunami* and *Japan*. Another prominent example is the use of social networks during Hurricane Sandy. As published by the official *@twitter* account on November 2nd, 2012: *"people sent more than 20 million tweets about the storm between Oct 27 and Nov 1."* This in turn means that tweets have become an interesting source of information for various people involved in crisis monitoring, such as journalists or crisis operators [6].

As data generated by citizens become more and more useful during emergencies, it is increasingly important to support the active tracking and analysis of these data. With WallTweet, we aim to contribute a tool that is useful and effective during real-time crisis monitoring. The tool relies on a geographic map of the monitored crisis event. Several visualization techniques are used to provide local detail in a global geographic context, in order to support situation awareness. All tweets are analyzed according to their semantic context. Figure 1 shows one of the visualizations we are currently developing: tweets are geographically included into clusters and semantically analyzed to identify the most relevant topic. In its current stage of development, WallTweet works on offline data, and as such is useful for understanding and analyzing – in retrospect – how emergencies have unfolded. Our test dataset contains about 500,000 tweets collected during one of the most critical days of the Hurricane Sandy crisis. Our goal is to first finalize the main visualizations, perform a user study, and then extend the tool to enable live updates during a crisis event.

INFORMATION VISUALIZATION FOR TWITTER

Twitter is a micro-blogging platform created and launched in 2006. Official statics published on June 30, 2015 on about.twitter.com/company, indicate that this social network has reached 316 millions active users per month and that 500 million tweets are sent each day. Tweets focus on different topics, including personal feelings, events of general interest, and daily news [5]. Shared information can be seen as an interesting source for discovering what is going on and what the opinions of involved witnesses are.

Considering the amount of generated data on Twitter, a question arises: is it possible to efficiently access and analyze tweets? Different visual analytic tools have been described in the literature, that take advantage of data collected from streams of information, like social networks or blogs.

The Visual Backchannel [3] is an innovative tool for monitoring micro-blogging platforms during large-scale events using different visualization techniques. Each technique aims at emphasizing a particular aspect of the data: a streamgraph for visualizing topic evolution over time; a helical graph of the most active users participating in the discussion; a list of tweets; a cloud of all published images.

Steed et al. [12] proposed a system called Matisse for automatically extracting emotions from the text messages and relating them with other measures, such as frequency of contained terms, time range and geographical coordinates. All these data are combined through three visualization techniques: a timeline, a streamgraph, and a map. With this system, it is possible to have a general understanding of how people feel about a specific event and from where they are posting.

Based on the same idea, Zimmerman and Vatrapu [16] built a prototype that combines information from different social media channels into six different dashboards. Three of them present general offline statistics about most relevant topics, sharing activities, *likes* from other users, and most active contributors in the networks. The other three dashboards are about the real-time evolution of the same information. Following the same direction, Hao et al. have focused on identifying customer opinions and possible useful patterns from tweets as influences on the market [4].

In order to understand the current experience of emergency managers and practitioners with social media, authors in literature have been contributing with the design of several user studies. This is the case of SensePlace2 of MacEachren et al. in [7] and ScatterBlogs of Thom et al. in [13]. In both contributions, participants involved in the evaluation agree on the relevance that social media have in today's crisis, suggesting also several issues to take into account for designing a tool for analyzing them, as for example the privacy or the adaptability to a specific situation.

All of the above systems are aimed at better supporting the understanding of how information propagates, and peoples' sentiments, when a large-scale event occurs. We are not aware of any visualization of tweets for high-resolution wall-sized displays. However, researchers have worked on the design of multi-surface interactive environments for crisis management centers, involving multiple devices such as tablets, smartphones, and both horizontal and vertical large displays [2]. In particular, the wall display's purpose is to give an overview of collected data from Twitter about the incident. The visualization relies on the Folding View technique, that distorts the information space depending on where the user's attention is directed. If users need more detailed information, a tablet or a smaller device is required.

THE KNOWLEDGE ECOSYSTEM

A Knowledge Ecosystem is defined by Thomson as a *"complex and many-faceted system of people, institutions, organizations, technologies and processes by which knowledge is created, interpreted, distributed, absorbed and utilized"* [14]. The Knowledge Ecosystem used within WallTweet is the result of a semantic similarity-based approach for analyzing text, already presented by Onorati and Diaz [10]. It consists of four different steps: (i) query Twitter for one or more keywords; (ii) perform a syntactic analysis of collected tweets for extracting nouns, where nouns are considered the most meaningful elements in a speech; (iii) filter extracted nouns, identifying the relevant ones by comparing their frequencies with a domain ontology about emergency; (iv) perform a semantic analysis of filtered terms, associating each one with a fixed category. Categories have a semantic value; they help in organizing the tweets depending on their correlation with *Emergency*, *Evacuation*, *Media*, *Hashtag*, *Place*, *Time* and *General*. All these data, including tweets, extracted terms, frequencies and categories, are visualized in WallTweet using different techniques, as described in the next section.

In this work, we have applied this mechanism for collecting information from the Hurricane Sandy dataset. While at the end of the hurricane crisis more than 20 million tweets were published, we currently work with a subset consisting of almost 500,000 tweets. These tweets are the result of querying keywords *hurricane* and *Sandy*, as well as hashtags *#hurricane* and *#sandy* during the first 24 hours of the hurricane hitting New York bay on October 29. During the semantic analysis, almost 24,000 nouns were extracted, successively filtered and reduced to 5,500.

WALLTWEET

Emergency operation centers usually work with a large display showing a map visualizing the current position of officers on duty, temperature, traffic information, etc. Inspired by this setup, we use a wall displays that shows a geographic map

Figure 2. The three proposed views: (a) *Global View*, (b) *Semantic View*, (c) *Time Sequence*.

of the hurricane twitter data. Our goal is to explore different options for helping operators in finding an answer to questions about the most affected areas, the number of people involved, the range of damages, or the effects of the rescue activities.

The Knowledge Ecosystem described above runs on the WILDER ultra-high-resolution wall, that consists of 75 narrow-bezel LED tiles (960x960 pixels each, 60ppi) laid out in a 15x5 matrix, 6 meters wide and 2 meters high for a total resolution of 14,400 × 4,800 pixels. WILDER is driven by a cluster of 10 computers, each equipped with high-end graphics cards, and a master workstation. The platform also features multiple input capabilities, including a multi-touch frame, real-time motion tracking, and handheld devices. WallTweet is built upon jBricks [11], a Java toolkit for rapidly prototyping multi-scale interfaces on cluster-driven wall displays.

Ultra-walls (short for *ultra-high-resolution wall displays* [9]) make it possible to visualize much larger volumes of data compared to earlier projector-based wall displays, whose pixel density is lower. Ultra-walls support the display of large datasets with a high level of detail while retaining context. They afford multi-scale interaction through physical navigation [1]: users can move from an overview of the data, to the fine details of a specific area simply by walking in front of the wall.

WallTweet offers three main views: *Global View*, *Semantic View* and *Time Sequence*. In the *Global View*, geo-located tweets are represented by geographical points over a map. Each point gets assigned a background color depending on the semantic category of the terms contained in the tweet (Figure 2-a). This view also includes a bubble chart representing the most relevant terms extracted from the dataset. Analyzing this view, we can see that the most discussed topics are related to the emergency description, and the majority of the tweets are distributed around big cities, like New York City, Baltimore, Washington and Boston. This can indicate densely populated areas, and areas that have been impacted most by the hurricane. Knowing where these areas are located is a crucial information for emergency operators in charge of making decisions about where to allocate resources.

The *Semantic View* combines two different visualizations (Figure 2-b): a geographical clustering and a tag cloud. The clustering consists of groups of tweets that are geographically close. To identify these clusters, we apply the *concave hull* algorithm [8], that associates each point with its neighbors at a distance that has been determined empirically. For each cluster, tweets are semantically analyzed based on the Knowledge Ecosystem described earlier, in order to identify the most relevant terms and use them as labels. As in tag clouds, the labels' font size depends on their relevance with respect to the defined Knowledge Ecosystem. This view thus gives an idea of the semantic distribution of terms with respect to their geographical position.

The last view, *Time Sequence*, is an animated visualization in which each tweet falls from the top of the display and is added to the map at the time of its posting on the social network. Clusters are updated once the tweet reaches its geographical position. Figure 2-c illustrates the visualization after new tweets have fallen on the map, showing that both the labels' font size and the clusters have changed. Each cluster has its own label (indicating the topic) and is painted with a specific color to make it easier to distinguish. If used in real-time, this dynamic visualization of how published content evolves over time can help emergency operators or media centers to get an idea of how people are reacting to the emergency, and taking these reactions into account to inform future decisions.

DEVELOPING VISUALIZATIONS FOR WALL DISPLAYS

The benefit of using a large surface with a ultra-high resolution is twofold. First, its large size enables multiple users to work together in a shared workspace, thus making the monitoring exercise and the collaboration among them easier. Second, as mentioned earlier, the very-high pixel density enables users to look at the data at different levels of detail simply by physically moving in front of the display. Developing multi-scale collaborative visualizations that take advantage of these properties is not trivial, but we have started to make progress. For example, in our Semantic View, clusters can be seen at a distance from the wall, indicating pockets of activity, while when stepping closer users can also read details about the most discussed topics of each cluster. The map can also be smoothly zoomed in and out, in which case clusters get recomputed as the distance used by the concave hull algorithm to compute point neighborhoods is adapted to the new zoom factor. Interacting with the display, it is also possible to read tweets and compare them with the performed semantic analysis.

Walltweet can be seen as part of a more complex ecology of devices aiming at surrounding users – who have to handle crisis situations – with data, helping them explore those data and extract meaningful information from them in an efficient and timely manner. We believe that using wall displays in such an ecology can significantly improve users' situation awareness, and we are very interested in further discussing the associated challenges in this workshop.

CONCLUSIONS AND FUTURE WORKS

Social networks and other messaging services have drawn the attention of researchers and practitioners in media and crisis management. In social networks like Twitter, people share their opinions and experiences, generating vast quantities of data about a wide range of topics, including real-time information on crises that impact them. Visualizing this data in order to better understand and use it for decision making is a challenging topic in visual analytics and information visualization.

In this paper, we introduced a prototype system for analyzing and visualizing tweets generated during a large-scale critical situation, in order to support the monitoring activities of emergency operation and media centers. The main contributions of our approach so far are the combination of the semantic analysis of tweets and the use of a ultra-high-resolution wall display for visualizing its results. We are still at an early stage of this on-going project, called WallTweet. So far it consists of three different views: a generic view showing the semantic categorization of tweets on the map; a semantic view with geographical clusters of tweets, tagged with the most relevant terms associated with them; a time sequence that simulates the real-time posting of tweets on the map, the definition and evolution of clusters, and the varying relevance of terms.

Future work will focus on two different directions. First, we are going to evaluate WallTweet and its visualizations from both a domain and a usability point of view. We are currently planning a usability test as well as an expert evaluation with emergency practitioners. Second, we are going to make the system run in real-time, integrating it in an ecology of devices for achieving better collaborative sense making and higher situation awareness: including additional data sources such as digital sensors for tracking information about, e.g., pressure, temperature or traffic, and providing users with more elaborate capabilities for interacting with the wall display using devices such as smartphones and tablets.

ACKNOWLEDGMENTS

This work is supported by the project emerCien grant funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-09687). Teresa Onorati was partially supported by a grant from Universidad Carlos III de Madrid.

REFERENCES

1. R. Ball, C. North, and D. Bowman. 2007. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proc. CHI*. ACM, 191–200.
2. S. Butscher, J. Müller, A. Weiler, R. Rädle, H. Reiterer, and M. H Scholl. 2013. Multi-user Twitter Analysis for Crisis Room Environments. In *3rd Crisis Management Technology Workshop on Collaborative Human Computer Interaction with Big Wall Displays*. 28–34.
3. M. Dork, D. Gruen, C. Williamson, and S. Carpendale. 2010. A Visual Backchannel for Large-Scale Events. *IEEE Trans. on Vis. and Comp. Graphics (TVCG)* 16, 6 (2010), 1129–1138.
4. M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Keim, L. E. Haug, and M. C. Hsu. 2011. Visual sentiment analysis on twitter data streams. In *Proc. VAST*. 277–278.
5. A. Java, X. Song, T. Finin, and B. Tseng. 2007. Why we twitter: understanding microblogging usage and communities. In *9th WebKDD workshop on Web mining and social network analysis*. ACM, 56–65.
6. B. R. Lindsay. 2010. Social Media and Disasters: Current Uses, Future Options and Policy Considerations. *Journal of Current Issues in Media & Telecommunications* 2 (2010), 287–297.
7. A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. 2011. SensePlace2: GeoTwitter analytics support for situational awareness. *Proc. VAST* (2011), 181–190.
8. A. Moreira and M. Y. Santos. 2006. Concave Hull: A k-Nearest Neighbours Approach for The Computation of The Region Occupied By A Set of Points. In *Conference on Computer Graphics Theory and Applications*. 61–68.
9. M. Nancel, E. Pietriga, O. Chapuis, and M. Beaudouin-Lafon. 2015. Mid-Air Pointing on Ultra-Walls. *ACM ToCHI* 22, 5 (2015), 1–62.
10. T. Onorati and P. Diaz. 2015. Semantic Visualization of Twitter Usage in Emergency and Crisis Situations. *Proc. ISCRAM (to appear)* (2015).
11. E. Pietriga, S. Huot, M. Nancel, and R. Primet. 2011. Rapid development of user interfaces on cluster-driven wall displays with jBricks. In *Proc. EICS*. ACM, 185–190.
12. C. A. Steed, J. Beaver, P. L. Bogen II, M. Drouhard, and J. Pyle. 2015. Text Stream Trend Analysis using Multiscale Visual Analytics with Applications to Social Media Systems. In *ACM IUI Workshop on Visual Text Analytics*.
13. D. Thom, R. Krger, U. Bechstedt, A. Platz, J. Zisgen, B. Volland, and T. Ertl. 2015. Can twitter really save your life? A case study of visual social media analytics for situation awareness. In *IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 183–190.
14. A. J. Thomson, B. E. Callan, and J. J. Dennis. 2007. A knowledge ecosystem perspective on development of web-based technologies in support of sustainable forestry. *Computers and Electronics in Agriculture* 59, 1–2 (2007), 21–30.
15. O. Westlund. 2013. Mobile News. *Digital Journalism* 1, 1 (Feb. 2013), 6–26.
16. C. Zimmerman and R. Vatrapu. 2015. The Social Newsroom: Visual Analytics for Social Business Intelligence Tool Prototype Demo. *New Horizons in Design Science.* LNCS 9073 (2015), 386–390.